Toward Efficient and Secure Public Auditing for Dynamic Big Data Storage on Cloud

by

Chang Liu

B. Sci. (Shandong University)

M. Eng. (Shandong University)

A thesis submitted to

Faculty of Engineering and Information Technology

University of Technology, Sydney

for the degree of

Doctor of Philosophy

December 2014

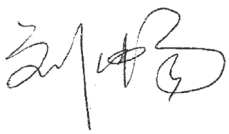
To my family and friends

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Chang Liu



21 July 2015

Acknowledgements

I sincerely express my deepest gratitude to my primary supervisor, A/Prof. Jinjun Chen, for his seasoned supervision and continuous encouragement throughout my PhD study. Prof. Chen was a consistent inspiration and taught me a great deal about how to become a good researcher and a good person. I also thank my associate supervisors Dr. Rajiv Ranjan, Prof. Yun Yang and Prof. Chengfei Liu for their whole-hearted supervision and continuous support of my study.

I thank the Australian Research Council (ARC), Australia Commonwealth Scientific and Industrial Research Organization (CSIRO) and the University of Technology, Sydney (UTS) for offering me a full research scholarship throughout my doctoral program. I also thank the CSIRO Digital Productivity Flagship (DPF) and the Research Committee of the UTS Faculty of Engineering and Information Technology (FEIT) for research publication funding support and for providing me with financial support to attend conferences.

My thanks also goes to staff members and researchers at UTS FEIT, CSIRO DPF, Swinburne University of Technology and Nanjing University for their help, suggestions, friendship and encouragement, particularly: Prof. Igor Hawryszkiewycz, A/Prof. Maolin Huang, Indrawati Nataatmadja, Xuyun Zhang, Chi Yang, Miranda Qian Zhang, Adrian Johannes, Nazanin Borhan, A. Ali, Xiao Liu, Dong Yuan, Qiang He, Rui Zhou, Jianxin Li, Minyi Li, Wei Dong, Dahai Cao, Miao Du, Feifei Chen, Prof. Wanchun Dou, Wenmin Lin.

Last but not least, I am deeply grateful to my parents Hongbo Liu and Li Zheng for raising me, teaching me to be a good person, and supporting my studies abroad. Sadly, my dear father Hongbo passed away in 2013 during my PhD study. May he rest in peace.

Abstract

Cloud and Big Data are two of the most attractive ICT research topics that have emerged in recent years. Requirements of big data processing are now everywhere, while the pay-as-you-go model of cloud systems is especially cost efficient in terms of processing big data applications. However, there are still concerns that hinder the proliferation of cloud, and data security/privacy is a top concern for data owners wishing to migrate their applications into the cloud environment. Compared to users of conventional systems, cloud users need to surrender the local control of their data to cloud servers. Another challenge for big data is the data dynamism which exists in most big data applications. Due to the frequent updates, efficiency becomes a major issue in data management. As security always brings compromises in efficiency, it is difficult but nonetheless important to investigate how to efficiently address security challenges over dynamic cloud data.

Data integrity is an essential aspect of data security. Except for server-side integrity protection mechanisms, verification from a third-party auditor is of equal importance because this enables users to verify the integrity of their data through the auditors at any user-chosen timeslot. This type of verification is also named 'public auditing' of data. Existing public auditing schemes allow the integrity of a dataset stored in cloud to be externally verified without retrieval of the whole original dataset. However, in practice, there are many challenges that hinder the application of such schemes. To name a few of these, first, the server still has to aggregate a proof with the cloud controller from data blocks that are distributedly stored and processed on cloud instances and this means that encryption and transfer of these data within the cloud will become time-consuming. Second, security flaws exist in the current designs. The verification processes are insecure against various attacks and this leads to concerns about deploying these schemes in practice. Third, when the dataset is large, auditing of dynamic data becomes costly in terms of communication and storage. This is especially the case for a large number of small data updates and data updates on

multi-replica cloud data storage.

In this thesis, the research problem of dynamic public data auditing in cloud is systematically investigated. After analysing the research problems, we systematically address the problems regarding secure and efficient public auditing of dynamic big data in cloud by developing, testing and publishing a series of security schemes and algorithms for secure and efficient public auditing of dynamic big data storage on cloud. Specifically, our work focuses on the following aspects: cloud internal authenticated key exchange, authorisation on third-party auditor, fine-grained update support, index verification, and efficient multi-replica public auditing of dynamic data. To the best of our knowledge, this thesis presents the first series of work to systematically analysis and to address this research problem. Experimental results and analyses show that the solutions that are presented in this thesis are suitable for auditing dynamic big data storage on cloud. Furthermore, our solutions represent significant improvements in cloud efficiency and security.

The Author's Publications

Book Chapters:

1. **C. Liu**, R. Ranjan, X. Zhang, C. Yang and J. Chen, *A Big Picture of Integrity Verification of Big Data in Cloud Computing*, Handbook on Data Centers (Book), Springer, in press, 2014.

2. X. Zhang, **C. Liu**, S. Nepal, C. Yang and J. Chen, *Privacy Preservation over Big Data in Cloud Systems*, Security, Privacy and Trust in Cloud Systems (Book), Springer, in press, ISBN: 978-3-642-38585-8, 2013.

Journals:

3. **C. Liu**, R. Ranjan, C. Yang, X. Zhang, L. Wang and J. Chen, *MuR-DPA: Top-down Levelled Multi-replica Merkle Hash Tree Based Secure Public Auditing for Dynamic Big Data Storage on Cloud*, IEEE Transactions on Computers, accepted on 27 October, 2014

4. **C. Liu**, N. Beaugeard, C. Yang, X. Zhang and J. Chen, *HKE-BC: Hierarchical Key Exchange for Secure Scheduling and Auditing of Big Data in Cloud Computing*, Concurrency and Computation: Practice and Experience, accepted on 3 October, 2014

5. X. Zhang, W. Dou, J. Pei, S. Nepal, C. Yang, **C. Liu** and J. Chen, *Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud*, IEEE Transactions on Computers, accepted on 18 August, 2014.

6. **C. Liu,** C. Yang, X. Zhang and J. Chen, *External Integrity Verification for Outsourced Big Data in Cloud and IoT: A Big Picture*, Future Generation

Computer Systems (FGCS), Elsevier, to appear, accepted on 16 August, 2014. doi: 10.1016/j.future.2014.08.007

7. W. Lin, W. Dou, Z. Zhou and **C. Liu**, *A Cloud-based Framework for Home-diagnosis Service over Big Medical Data*, Journal of Systems and Software (JSS), to appear, accepted on 22 May, 2013. (ERA Rank A)

8. C. Yang, **C. Liu**, X. Zhang, S. Nepal and J. Chen, *A Time Efficient Approach for Detecting Errors in Big Sensor Data on Cloud*, IEEE Transactions on Parallel and Distributed Systems (TPDS), to appear, accepted on 7 December, 2013. (ERA Rank A*)

9. **C. Liu**, J. Chen, L. T. Yang, X. Zhang, C. Yang, R. Ranjan and K. Ramamohanarao, *Authorized Public Auditing of Dynamic Big Data Storage on Cloud with Efficient Verifiable Fine-grained Updates*, IEEE Transactions on Parallel and Distributed Systems (TPDS), vol. 25, no. 9, pp. 2234-2244, 2014. (ERA Rank A*)

10. X. Zhang, **C. Liu**, S. Nepal, C. Yang, W. Dou and J. Chen, *A Hybrid Approach for Scalable Sub-Tree Anonymization over Big Data using MapReduce on Cloud*, Journal of Computer and System Sciences (JCSS), vol. 80, no. 5, pp. 1008–1020, 2014. (ERA Rank A*)

11. C. Yang, X. Zhang, C. Zhong, **C. Liu**, J. Pei, K. Ramamohanarao and J. Chen, *A Spatiotemporal Compression based Approach for Efficient Big Data Processing on Cloud*, Journal of Computer and System Sciences (JCSS), to appear, 2013. (ERA Rank A*)

12. C. Yang, X. Zhang, **C. Liu**, J. Pei, K. Ramamohanarao and J. Chen, *A Spatiotemporal Compression based Approach for Efficient Big Data Processing on Cloud*, Journal of Computer and System Sciences (JCSS), to appear, 2013. (ERA Rank A*)

13. X. Zhang, **C. Liu**, S. Nepal, C. Yang, W. Dou and J. Chen, SaC-FRAPP: *A Scalable and Cost-effective Framework for Privacy Preservation over Big Data on Cloud*, Concurrency and Computation: Practice and Experience (CCPE), vol. 25, no. 18, pp. 2561-2576, 2013. (ERA Rank A)

14. X. Zhang, L. T. Yang, **C. Liu** and J. Chen, *A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization using MapReduce on Cloud*, IEEE Transactions on Parallel and Distributed Systems (TPDS), 25(2): 363-373, 2014. (ERA Rank A*)

15. X. Zhang, **C. Liu**, S. Nepal and J. Chen, *An Efficient Quasi-identifier Index based Approach for Privacy Preservation over Incremental Data Sets on Cloud*, Journal of Computer and System Sciences (JCSS), 79(5): 542-555, 2013. (ERA Rank A*)

16. X. Zhang, **C. Liu**, S. Nepal, S. Panley and J. Chen, *A Privacy Leakage Upper-bound Constraint based Approach for Cost-effective Privacy Preserving of Intermediate Datasets in Cloud*, IEEE Transactions on Parallel and Distributed Systems (TPDS), vol. 24, no. 6, pp. 1192-1202, 2013. (ERA Rank A*)

17. **C. Liu**, X. Zhang, C. Yang and J. Chen, *CCBKE - Session Key Negotiation for Fast and Secure Scheduling of Scientific Applications in Cloud Computing*, Future Generation Computer Systems (FGCS), Elsevier, vol. 29, no. 5, pp. 1300-1308, 2013. (ERA Rank A)

Conferences:

18. **C. Liu**, R. Ranjan, X. Zhang, C. Yang, D. Georgakopoulos and J. Chen, *Public Auditing for Big Data Storage in Cloud Computing -- A Survey*, in Proc. The 16th IEEE International Conference on Computational Science and Engineering (CSE 2013), pp. 1128-1135, December, 2013, Sydney, Australia.

19. C. Yang, **C. Liu**, X. Zhang, S. Nepal and J. Chen, *Querying Streaming XML Big Data with Multiple Filters on Cloud*, in Proc. The 2nd International Conference on Big Data and Engineering (BDSE 2013), pp. 1121-1127, December, 2013, Sydney, Australia.

20. X. Zhang, C. Yang, S. Nepal, **C. Liu**, W. Dou and J. Chen, *A MapReduce Based Approach of Scalable Multidimensional Anonymization for Big Data Privacy Preservation on Cloud*, in Proc. 3rd International Conference on Cloud and Green Computing (CGC 2013), pp: 105 - 112, September, 2013, Karlsurhe, Germany.

21. **C. Liu**, X. Zhang, C. Liu, Y. Yang, R. Ranjan, D. Georgakopoulos and J. Chen, *An Iterative Hierarchical Key Exchange Scheme for Secure Scheduling of Big Data Applications in Cloud Computing*, in Proc. 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom2013), pp: 9-16, July, 2013, Melbourne, Australia. (ERA Rank A)

22. X. Zhang, **C. Liu**, S. Nepal, C. Yang, W. Dou and J. Chen, *Combining Top-Down and Bottom-Up: Scalable Sub-Tree Anonymization over Big Data using MapReduce on Cloud*, in Proc. 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom2013), pp: 501-508, July, 2013, Melbourne, Australia. (ERA Rank A)

23. X. Zhang, **C. Liu**, S. Nepal, W. Dou and J. Chen, *Privacy-preserving Layer over MapReduce on Cloud*, in Proc. 2nd International Conference on Cloud and Green Computing (CGC 2012), pp: 304-310, November, 2012, Xiangtan, China.

24. G. Zhang, Y. Yang, X. Zhang, **C. Liu** and J. Chen, *Key Research Issues for Privacy Protection and Preservation in Cloud Computing*, in Proc. 2nd International Conference on Cloud and Green Computing (CGC 2012), pp: 304-310, November, 2012, Xiangtan, China.

25. G. Zhang, Y. Yang, X. Zhang, **C. Liu** and J. Chen, *An Association Probability based Noise Generation Strategy for Privacy Protection in Cloud Computing*, in Proc. 10th International Conference on Service Oriented Computing (ICSOC2012), November, 2012, Shanghai, China. (ERA Rank A)

26. **C. Liu**, X. Zhang, J. Chen and C. Yang, *An Authenticated Key Exchange Scheme for Efficient Security-Aware Scheduling of Scientific Applications in Cloud Computing*, In Proc. 9th IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC2011). pp: 372-379, December, 2011, Sydney, Australia.

27. X. Zhang, **C. Liu**, J. Chen and W. Dou, *An Upper-Bound Control Approach for Cost-Effective Privacy Protection of Intermediate Dataset Storage in Cloud*, In Proc. 9th IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC2011). pp: 518-525, December, 2011, Sydney, Australia.

28. C. Yang, K. Ren, Z. Yang, P. Gong and **C. Liu**, *A CSMA-based Approach for Detecting Composite Data Aggregate Events with Collaborative Sensors in WSN*, In Proc. 15th International Conference on Computer Supported Cooperative Work in Design (CSCWD2011). pp: 489-496, June, 2011, Lausanne, Switzerland.

29. C. Yang, Z. Yang, K. Ren and **C. Liu**, *Transmission Reduction based on Order Compression of Compound Aggregate Data over Wireless Sensor Networks*, In Proc. 6th International Conference on Pervasive Computing and Applications (ICPCA2011). October, 2011, Port Elizabeth, South Africa.

Table of Contents

Figures ..... **xiv**

Tables ..... **xvii**

Chapter 1 Introduction **1**

1.1 Big Data and Cloud Computing 1

1.2 Security and Privacy Concerns in Cloud 3

1.3 Public Auditing of Dynamic Cloud Data 3

1.4 Thesis Overview 7

Chapter 2 Literature Review **10**

2.1 Security and Privacy Research in Cloud and Big Data 10

2.2 Integrity Verification and Public Auditing 13

2.3 Authenticated Key Exchange in Cloud 17

Chapter 3 Background, Problem Analysis and Framework **20**

3.1 Preliminaries 20

3.1.1 Diffie-Hellman Key Exchange 20

3.1.2 RSA Signature 21

3.1.3 Bilinear Pairing and BLS Signature 22

3.1.4 Authenticated Data Structures 22

3.2 Motivating Examples and Research Framework 24

3.2.1 Motivating Examples 24

3.2.2 Research Problems with Public Auditing of Cloud Data -- Lifecycle and Framework 28

3.3 Representative Public Auditing Schemes In Sketch 33

3.3.1 PDP 33

3.3.2 Compact POR 34

3.3.3 DPDP 35

3.3.4 Public Auditing of Dynamic Data 36

3.4 Detailed Analysis of Research problems 37

3.4.1 Authenticated Key Exchange in Cloud 37

3.4.2 Public Auditing of Verifiable Fine-grained Updates 40
3.4.3 Multi-replica Big Data in Cloud 42
3.4.4 Security of Public Auditing Schemes 43

Chapter 4 Authenticated Key Exchange Schemes in Cloud 47
4.1 CCBKE: Cloud Computing Background Key Exchange 47
4.1.1 System setup 47
4.1.2 Key Exchange 48
4.1.3 Rekeying 50
4.2 HKE-BC: Hierarchical Key Exchange for Big data in Cloud 51
4.2.1 System Setup 52
4.2.2 Key Exchange 52
4.3 Security and Efficiency Analysis 58
4.3.1 Security Proofs 58
4.3.2 Perfect Forward Secrecy 61
4.3.3 Efficiency Analysis for HKE-BC 62

Chapter 5 FU-DPA: Public Auditing for Dynamic Data with Fine-grained Updates 65
5.1 Introduction 65
5.2 Preliminaries 67
5.2.1 Bilinear Pairing 67
5.2.2 Weighted Merkle Hash Tree 67
5.3 Framework and Definitions for Supporting Fine-grained Updates 68
5.4 The Proposed Scheme 70
5.4.1 First Scheme 70
5.4.2 Analysis on Fine-grained Dynamic Data Updates 76
5.4.3 Further Modification for Better Support of Small Updates 83
5.4.4 Further Discussions 84
5.5 Security and Efficiency Analysis 85
5.5.1 Security Analysis 85
5.5.2 Efficiency Analysis 89

Chapter 6 MuR-DPA: Secure Public Auditing for Dynamic Multi-replica Big

Data Storage on Cloud 93
6.1 Introduction 93
6.2 Preliminaries 95
6.2.1 Bilinear Pairing 95
6.2.2 Rank-based Multi-Replica Merkle Hash Tree 95
6.3 Verification of All Replicas at Once 98
6.4 Efficient Verifiable Updates on Multi-replica Cloud Data 100
6.5 Discussions and Extensions 104
6.6 Security and Efficiency Analysis 106
6.6.1 Verifiable Multi-Replica Updates 106
6.6.2 All-at-once Multi-Replica Verification 108

Chapter 7 Experimental Results and Evaluations 111
7.1 Qualitative Comparison of Public Auditing Schemes 111
7.2 Experimental Environment 111
7.3 Experimental Results for Key Exchange Schemes 113
7.3.1 Comparison of Key Exchange schemes 113
7.3.2 Efficiency improvements of CCBKE and HKE-BC 116
7.4 Experimental Results for FU-DPA 120
7.5 Experimental Results for MuR-DPA 124

Chapter 8 Conclusions and Future Work 131
8.1 Conclusions 131
8.2 Future Work 131
8.2.1 Aspects for Measurements and Improvements 131
8.2.2 Future Research Problems 134

Bibliography 138

Appendix A Acronyms 148

Appendix B Notation Index 150

Figures

Figure 1-1 Thesis structure. 7

Figure 3-1 ADS examples: MHT and RASL. 24

Figure 3-2 Participating parties in public auditing of cloud data. 29

Figure 3-3 Integrity verification for outsourced data -- a framework 30

Figure 3-4 Integrity verification for outsourced data -- the lifecycle. 31

Figure 3-5 An example of hybrid cloud structures. 41

Figure 4-1 Process of HKE-BC Phase1. 52

Figure 4-2 Process of HKE-BC Phase2. 56

Figure 5-1 An example of a weighted Merkle hash tree (WMHT). 68

Figure 5-2 Verifiable *PM*-typed data update in FU-DPA. 74

Figure 5-3 Challenge, proof generation and verification in FU-DPA. 77

Figure 5-4 The algorithm to find a block in *F* with a given offset *o*. 78

Figure 5-5 Example: fine-grained insertion. 80

Figure 5-6 Example: fine-grained deletion. 82

Figure 5-7 Example: fine-grained modification. 83

Figure 5-8 Verifiable *PM*-typed data update in modified (final) FU-DPA. 85

Figure 6-1 An example of RMR-MHT 97

Figure 6-2 Public auditing of all replicas at once..... 101

Figure 6-3 Update examples to RMR-MHT 103

Figure 6-4 Dynamic data update and verification..... 105

Figure 7-1 U-Cloud environment..... 114

Figure 7-2 Structures of two cloud instantiations of U-Cloud. 118

Figure 7-3 Time efficiency of HKE-BC, CCBKE and IKE in the two cloud instantiations..... 119

Figure 7-4. Efficiency advantage of HKE-BC 120

Figure 7-5 Auditing communication overhead in FU-DPA for different block size. 121

Figure 7-6 FU-DPA: Comparison of storage overhead..... 121

Figure 7-7 FU-DPA: Comparison of storage overhead (continued). 122

Figure 7-8 FU-DPA: Reduction of communication overhead..... 122

Figure 7-9 MuR-DPA: Length of server response for one verifiable modification/insertion of one block. 126

Figure 7-10 MuR-DPA: Total communication for one verifiable update. 127

Figure 7-11 MuR-DPA: Extra storage overhead at server side for support of public auditability and data dynamics..... 128

Figure 7-12 MuR-DPA: Total communication overhead for auditing of all

replicas. 129

Figure 7-13 MuR-DPA: Communication for auditing of 1 chosen replica for a dataset with 1, 4 and 8 total replicas with different *s* value. 130

Tables

Table 7-1 Comparison of public auditing schemes - to be continued. 112

Table 7-2 Continued - comparison of public auditing schemes. 113

Table 7-3 Time consumption comparisons of IKE and AES encryption on CLC. 116

Table 7-4 Time consumption comparisons of IKE and Salsa encryption on CLC. 116

Table 7-5 Price of dynamism. 125

Table 8-1 Future Work..... 132